Two different approaches for attacking BOWS*

Pedro Comesaña and Fernando Pérez-González

Dept. Teoría de la Señal y Comunicaciones. ETSI Telecom., Universidad de Vigo, 36310 Vigo, Spain

ABSTRACT

From December 15, 2005 to June 15, 2006 the watermarking community was challenged to remove the watermark from 3 different 512×512 watermarked images while maximizing the *Peak Signal to Noise Ratio* (PSNR) measured by comparing the watermarked signals with their attacked counterparts. This challenge, which bore the inviting name of *Break Our Watermarking System* (BOWS),[1] and was part of the activities of the European Network of Excellence ECRYPT, had as its main objective to enlarge the current knowledge on attacks to watermarking systems; in this sense, BOWS was not aimed at checking the vulnerability of the specific chosen watermarking scheme against attacks, but to inquire in the different strategies the attackers would follow to achieve their target.

In this paper the main results obtained by the authors when attacking the BOWS system are introduced. Mainly, the strategies followed can be divided into two different approaches: *blind sensitivity attacks* and *exhaustive search of the secret key.*

1. INTRODUCTION

In the last decade, with the spreading of the Internet as an impressive communication tool and the appearance of advanced editing tools which can be used by almost any non-trained user, the need of new technical solutions to the problems of copyright protection, authentication, fingerprinting or annotation of digital contents have been raised. Digital watermarking has been widely recognized as a potentially powerful instrument against piracy, illegal modification, or improper use of contents. Nevertheless, experience has shown that the challenge of designing a watermarking method robust against an active attacker is extremely difficult. Even without considering geometrical attacks (which can be regarded to as some of the most harmful attacks against watermarking techniques), the range of strategies an attacker could envisage to remove the watermark from a watermarked content is virtually more diverse as the attackers theirselves.

We believe that challenging the watermarking community (and the public in general) to break a certain watermarking system is valuable for a number of reasons: 1) the contest serves to pinpoint the weaknesses of state-of-the-art methods, and likely, promote new research aimed at improving those methods; 2) the inherent applicability of the attacks serves as a benchmark to test results developed under more theoretical conditions; 3) the existence of independent attackers acts in a way as a "Monte Carlo" testing of the algorithms.

The design of *good (and new) attacks* is one of the main motivations of the *Break Our Watermarking System* (BOWS) challenge. This *contest* consists of removing the watermark from 3 watermarked signals, trying to maximize the Peak Signal to Noise Ratio (PSNR)[†], a squared error distortion measure. The fact of choosing a Mean Square Error (MSE) measure could be criticized, as it does not really reflect the impact of the attack on the semantics of the signal, but the lack of a universally recognized perceptual distortion measure also makes

*This work was partially funded by *Xunta de Galicia* under projects PGIDT04 TIC322013PR and PGIDT04 PXIC32202PM; MEC project DIPSTICK, reference TEC2004-02551/TCM; FIS project IM3, reference G03/185 and European Comission through the IST Programme under Contract IST-2002-507932 ECRYPT. ECRYPT disclaimer: The information in this paper is provided as is, and no guarantee or warranty is given or implied that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

Further author information: (Corresponding author: P.C.)
P.C.: E-mail: pcomesan@gts.tsc.uvigo.es, Telephone: +34 986 812683
F.P.-G.: E-mail: fperez@gts.tsc.uvigo.es, Telephone: +34 986 812124

[†]Remember that for 8 bits signals $\mathrm{PSNR}(\mathbf{x},\mathbf{y}) \triangleq 10\log_{10}\left[\frac{L\cdot 255^2}{||\mathbf{x}-\mathbf{y}||^2}\right]$, where L is the length of the compared signals.

(a) (b) (c)

Figure 1. The three watermarked signals provided by the organization.

difficult the choice of a non-MSE based measure. On the other hand, with the use of an MSE measure, the use of geometrical attacks for removing the watermark is precluded; these attacks, whose perceptual impact may be quite reduced, are known to be extremely harmful to the performance of most watermarking methods, as they often cause a loss in the synchronization of the watermark.

Furthermore, given that the organizers were also interested in investigating if the knowledge of the watermarking scheme could be useful for designing an attacking strategy, BOWS was divided in two different stages: for the first three months, just three 512×512 watermarked images and a binary detector were provided; later, and for the next 3 months, the watermarking method was made public.

With this framework in mind we tried to remove the watermark from the provided images in two different circumstances:

1. The attacker completely lacks any knowledge of the used watermarking method, and only has access to a detector, that he feeds with an image, and provides a binary output. This situation corresponds to the first stage of BOWS challenge.

2. The attacker knows all the details about the watermarking scheme, except for a secret parameter, the *the secret key*, which is only shared by embedder and detector.

For the first case we used a blind sensitivity attack previously published,[2] whereas for the second one we followed a strategy based on the exhaustive search on the space of the secret key. Detailed information about both methods is given in Sections 2 and 3 respectively, and their results are compared in Section 4.

2. BLIND SENSITIVITY ATTACKS

In a first stage of the contest the watermarking scheme used was not known by the attackers. In such case, a *blind* attack (meaning that it assumes no knowledge by the attacker of the watermarking technique) had to be used; in order to have a systematic algorithm, we chose the so-called *Blind Newton Sensitivity Attack* (BNSA),[2] which has shown to be effective against a wide range of watermarking methods. The characteristics of this algorithm which make it suitable for this kind of scenario are:

- It does not require any knowledge about the detection function.
- It just needs to know the binary output of the detection function for a given input (not the actual value of the detection function).
- It can be highly paralleled, in such a way that several attackers can work together, each of them using a different detector (or a set of them).

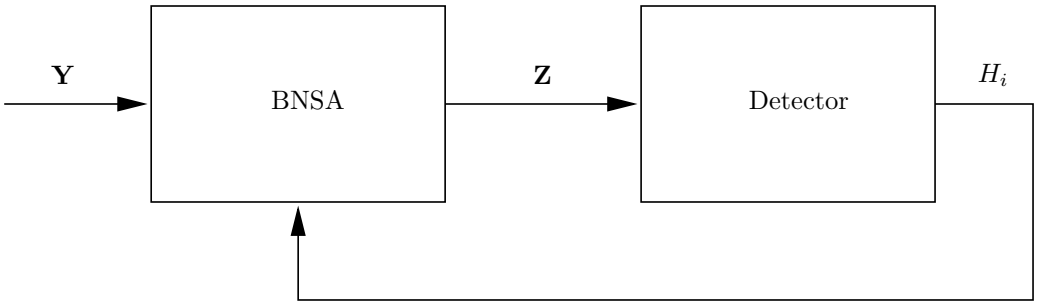

Figure 2. Block diagram of the BNSA.

- A full iteration of the algorithm is not necessary for obtaining relatively good results.

Furthermore, as we will explain later, it is not required to compute exactly the Hessian matrix and the gradient in order to be able to produce a descent direction of the target function. In fact, we will show that it is enough to compute just some of the components of the gradient vector in order to obtain signals where the watermark is already removed, but the quality of the attacked signal it is still really high. A basic block diagram of BNSA, depicting its iterative nature, is plotted in Fig. 2.

2.1. Blind Newton Sensitivity Attack

The BNSA[2] is based on formalizing the target of the attacker as

$$\arg \min_{\mathbf{t}:g(\mathbf{y}+\mathbf{t})\leq\eta} d_{\mathbf{y}}(\mathbf{t}),$$

where $d_{\mathbf{y}}(\mathbf{t})$ quantifies the distortion introduced by $\mathbf{t}$ on the watermarked signal $\mathbf{y}$, $g(\cdot)$ is the detection function, and η is a threshold which determines the detection region, in such a way that if $g(\mathbf{y}) > \eta$, the detector will determine that the watermark is present, and in any other case it will decide that there is not watermark in the received signal. Given that in our problem the objective is to maximize the PSNR, or equivalently to minimize the MSE of the distorting vector, we can see that in this particular scenario $d_{\mathbf{y}}(\mathbf{t}) = ||\mathbf{t}||^2$.

Therefore, the BNSA tries to iteratively solve this problem by using a surjection $h_{\mathbf{y}}$ with some specific characteristics, yielding an update of the algorithm with the generic form

$$\mathbf{s}_{k+1} = \mathbf{s}_k - \xi_k \cdot \mathbf{B}^{-1} \cdot \hat{\nabla}(d^{\star}_{\mathbf{y}} \circ h_{\mathbf{y}})(\mathbf{s}_k),$$

where $\hat{\nabla}(d^{\star}_{\mathbf{y}} \circ h_{\mathbf{y}})(\mathbf{s}_k)$ is the estimate of the gradient of $(d^{\star}_{\mathbf{y}} \circ h_{\mathbf{y}})(\mathbf{s}_k)$, and its i-th component is computed as

$$[\hat{\nabla}(d^{\star}_{\mathbf{y}} \circ h_{\mathbf{y}})(\mathbf{s}_k)]_i = \frac{(d^{\star}_{\mathbf{y}} \circ h_{\mathbf{y}})(\mathbf{s} + \delta\mathbf{e}_i) - (d^{\star}_{\mathbf{y}} \circ h_{\mathbf{y}})(\mathbf{s})}{\delta}.$$

This computation requires also to estimate $h_{\mathbf{y}}(\cdot)$, which is not known by the attacker. The proposed strategy is to use $h_{\mathbf{y}}(\mathbf{s}) = \alpha^* \cdot \mathbf{s}$, where $\alpha^* = \arg\min_{\alpha\in\mathbb{R}:g(\mathbf{y}+\alpha\mathbf{s})=\eta} |\alpha|$, computed using a bisection algorithm.

Concerning the matrix $\mathbf{B}$, different possibilities can be considered; probably the best choice would be to use an approximation of the Hessian, but due to computing limitations we have preferred to use $\mathbf{B} = \mathbf{I}_{L\times L}$, i.e. the identity matrix of size L. For a small enough ξ_k this choice of $\mathbf{B}$ guarantees a decrease of the target function as long as $(\hat{\nabla}(d^{\star}_{\mathbf{y}} \circ h_{\mathbf{y}})(\mathbf{s}_k))^T \cdot \nabla(d^{\star}_{\mathbf{y}} \circ h_{\mathbf{y}})(\mathbf{s}_k) > 0$, where the last condition is based on the Taylor series expansion of the objective function. In order to verify this condition, the attacker does not need to compute all the components of the estimate of the gradient; instead it is enough to calculate a small number of them and set the remaining to 0. Obviously, better results will be obtained when all the components were available, but the former strategy allows the attacker to stop the algorithm whenever he has obtained a suboptimal solution

which yields a satisfactory (following his quality criterion) attacked image; of course, adopting such strategy will reduce the computational cost of his attack.

Finally, the computation of the step-length[3, 4] of the k-th iteration ξ_k was performed following Armijo's rule, due to its simplicity.

2.2. Pseudocode

For the sake of clarity, next we give a pseudo-code description of the used implementation of the BNSA which has been debugged after close interaction with Prof. Barni's group in the University of Siena.[5] In the following description we assume that the provided watermarked signal $\mathbf{y}$ is arranged as a vector.

1. Generate an i.i.d. zero-mean Gaussian random vector $\mathbf{t}$, with the same size as the watermarked signal $\mathbf{y}$ and variance $\sigma_T^2 = 10^{-4}$.
2. Compute the minimum-normed scaling factor β of the vector $\mathbf{t}$ which is necessary for obtaining a non-watermarked signal when $\beta \cdot \mathbf{t}$ is added to $\mathbf{y}$. The squared Euclidean norm of the vector $\beta \cdot \mathbf{t}$ is denoted by γ_{start}.
3. For each component of the vector $\mathbf{t}$:
   (a) Slightly modify the i-th component of $\mathbf{t}$, obtaining $\mathbf{t}_i = \mathbf{t} + \epsilon_2 \cdot \mathbf{e}_i$, where $\mathbf{e}_i$ is the i-th vector of the canonical basis and $\epsilon_2 = 10^{-3}$.
   (b) Compute the minimum-normed scaling factor β of the vector $\mathbf{t}_i$ which is necessary for obtaining a non-watermarked signal when $\beta \cdot \mathbf{t}_i$ is added to $\mathbf{y}$. The squared Euclidean norm of the vector $\beta \cdot \mathbf{t}_i$ is denoted by γ_i.
4. Estimate the gradient of the squared Euclidean norm of the vector necessary for obtaining a non-watermarked signal, when the vector $\mathbf{t}$ is considered. The i-th component of the gradient is estimated as $\hat{\nabla}[i] = (\gamma_i - \gamma_{\text{start}})/\epsilon_2$.
5. Look for a step-size providing a decrease in the objective function:
   (a) $\xi = 10$.
   (b) $\mathbf{t}_{\text{new}} = \mathbf{t} - \xi \cdot \hat{\nabla}$.
   (c) Compute the minimum-normed scaling factor β of the vector $\mathbf{t}_{\text{new}}$ which is necessary for obtaining a non-watermarked signal when $\beta \cdot \mathbf{t}_{\text{new}}$ is added to $\mathbf{y}$. The squared Euclidean norm of the vector $\beta \cdot \mathbf{t}_{\text{new}}$ is denoted by γ_{after}.
   (d) While $\gamma_{\text{start}} < \gamma_{\text{after}}$:
      i. $\xi = 0.7 \cdot \xi$.
      ii. $\mathbf{t}_{\text{new}} = \mathbf{t} - \xi \cdot \hat{\nabla}$.
      iii. Compute the minimum-normed scaling factor β of the vector $\mathbf{t}_{\text{new}}$ which is necessary for obtaining a non-watermarked signal when $\beta \cdot \mathbf{t}_{\text{new}}$ is added to $\mathbf{y}$. The squared Euclidean norm of the vector $\beta \cdot \mathbf{t}_{\text{new}}$ is denoted by γ_{after}.
6. If the resulting signal $\mathbf{y}_2 = \mathbf{y} + \beta \cdot \mathbf{t}_{\text{new}}$ verifies the quality criteria established by the attacker, then $\mathbf{y}_2$ is the solution. Otherwise, the algorithm is iterated again from point 2 with $\mathbf{t} = \mathbf{t}_{\text{new}}$.

2.2.1. Computation of the minimum-normed scaling factor β of the vector $\mathbf{t}_0$ which is necessary for obtaining a non-watermarked signal when $\beta \cdot \mathbf{t}_0$ is added to $\mathbf{y}$

1. $\mathbf{t} = \mathbf{t}_0/||\mathbf{t}_0||$.
2. If either $\mathbf{y} + \mathbf{t}$ or $\mathbf{y} - \mathbf{t}$ is out of the detection region, then $\mathbf{v}_{\text{out1}} = \mathbf{t}$ and $\mathbf{v}_{\text{in1}} = \mathbf{0}$.
3. If both $\mathbf{y} + \mathbf{t}$ and $\mathbf{y} - \mathbf{t}$ are in the detection region:
   (a) While both $\mathbf{y} + \mathbf{t}$ and $\mathbf{y} - \mathbf{t}$ are in the detection region, $\mathbf{t} = 2 \cdot \mathbf{t}$.
   (b) $\mathbf{v}_{\text{out1}} = \mathbf{t}$ and $\mathbf{v}_{\text{in1}} = \mathbf{t}/2$.
4. If $\mathbf{y} + \mathbf{t}$ is out of the detection region:
   (a) $\mathbf{v}_{\text{out}} = \mathbf{v}_{\text{out1}}$ and $\mathbf{v}_{\text{in}} = \mathbf{v}_{\text{in1}}$.
   (b) While $||\mathbf{v}_{\text{out}} - \mathbf{v}_{\text{in}}|| > \epsilon_3 (= 10^{-3})$:
      i. $\mathbf{v}_{\text{middle}} = (\mathbf{v}_{\text{out}} + \mathbf{v}_{\text{in}})/2$.
      ii. If $\mathbf{y} + \mathbf{v}_{\text{middle}}$ is in the detection region, then $\mathbf{v}_{\text{in}} = \mathbf{v}_{\text{middle}}$; otherwise, $\mathbf{v}_{\text{out}} = \mathbf{v}_{\text{middle}}$.
   (c) $\mathbf{v}_+ = \mathbf{v}_{\text{out}}$
5. If $\mathbf{y} - \mathbf{t}$ is out of the detection region:
   (a) $\mathbf{v}_{\text{out}} = -\mathbf{v}_{\text{out1}}$ and $\mathbf{v}_{\text{in}} = -\mathbf{v}_{\text{in1}}$.
   (b) While $||\mathbf{v}_{\text{out}} - \mathbf{v}_{\text{in}}|| > \epsilon_3$:
      i. $\mathbf{v}_{\text{middle}} = (\mathbf{v}_{\text{out}} + \mathbf{v}_{\text{in}})/2$.
      ii. If $\mathbf{y} + \mathbf{v}_{\text{middle}}$ is in the detection region, then $\mathbf{v}_{\text{in}} = \mathbf{v}_{\text{middle}}$; otherwise, $\mathbf{v}_{\text{out}} = \mathbf{v}_{\text{middle}}$.
   (c) $\mathbf{v}_- = \mathbf{v}_{\text{out}}$
6. If $\mathbf{y} - \mathbf{t}$ is in the detection region, then $\mathbf{v} = \mathbf{v}_+$.
7. If $\mathbf{y} + \mathbf{t}$ is in the detection region, then $\mathbf{v} = \mathbf{v}_-$.
8. If both $\mathbf{y} + \mathbf{t}$ and $\mathbf{y} - \mathbf{t}$ are out of the detection region:
   - If $||\mathbf{v}_+|| < ||\mathbf{v}_-||$, then $\mathbf{v} = \mathbf{v}_+$; otherwise, $\mathbf{v} = \mathbf{v}_-$.
9. The minimum-normed scaling factor β of the vector $\mathbf{t}_0$ which is necessary for obtaining a non-watermarked signal when $\beta \cdot \mathbf{t}_0$ is added to $\mathbf{y}$ is given by the ratio between the value of any component of vectors $\mathbf{v}$ and $\mathbf{t}_0$, i.e. $\beta = \mathbf{v}[i]/\mathbf{t}_0[i]$, which is the same for any component.

2.3. Results

After one iteration of the BNSA performed in the pixel domain of the three provided images, the PSNR obtained for the first image is 56.3410 dB, for the second 56.9559 dB, and for the third one is 58.1586 dB. Furthermore, we think that is particularly interesting the fact that acceptably good results are achieved even when a full iteration of the BNSA is not performed. To illustrate, Fig. 4 shows the PSNR achieved versus the number of actually computed coefficients of the gradient. One can see that the PSNR obtained for 4096 computed coefficients of the gradient is already larger than 40 dB for all three images, meaning that a quite reduced amount of computations would be required for obtaining high quality contents where the watermark is removed. Furthermore, be aware that when the number of considered samples of the gradient is large enough (for values larger than approximately 2048), the PSNR grows almost linearly with the number of components of the gradient, meaning that the reduction in the attacking distortion (in dB) grows linearly with the available components of the gradient.

Finally, the attacker could try to learn something about the watermarking method. For example, although we have performed a BNSA in the pixel domain, a smart attacker could suspect that the images were not

(a) (b) (c)

Figure 3. The three signals obtained attacking the system with BNSA.

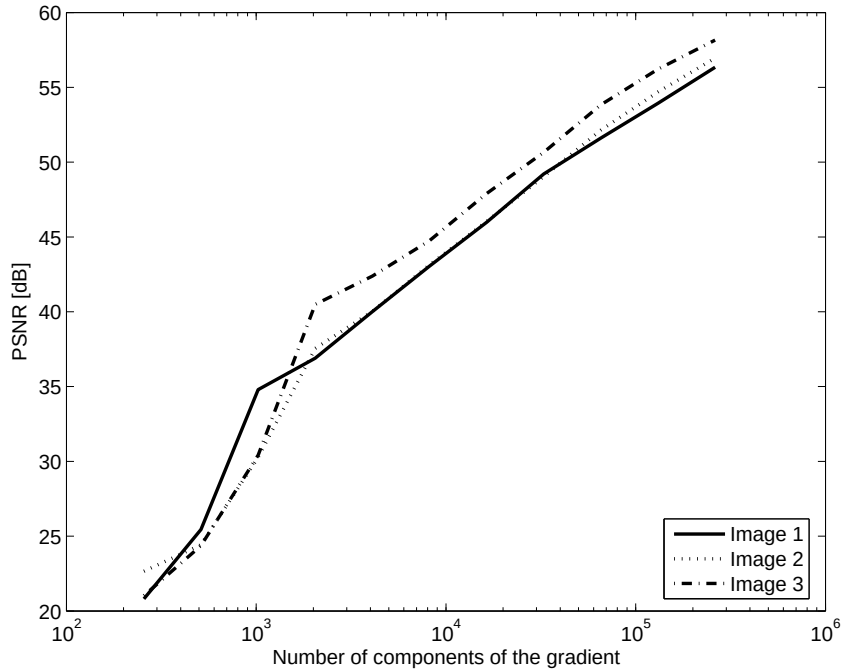

Figure 4. PSNR obtained as a function of the number of computed components of the gradient.

watermarked in all their frequency components, but just in a subset of them, as most of the algorithms in the literature use that kind of strategy. Therefore, after performing some tests the attacker could realize that the watermark was only embedded in 12 out of the 64 8×8 block DCT coefficients. Using that information, the attacker could speed up the attack more than 5 times, just by focusing on estimating the gradient of those block DCT coefficients that he knows (or assumes) that are being used by the watermarking method. Even though we did not pursue this line, we have determined the PSNR needed for removing the watermark from the provided images using the attacking vectors already computed with the BNSA in the spatial domain, and setting to 0 those 52 DCT components which are not used in each block. Remarkably, the results show an improvement on the obtained PSNR, yielding values of 57.5496, 57.8056 and 60.0081 dB, respectively.

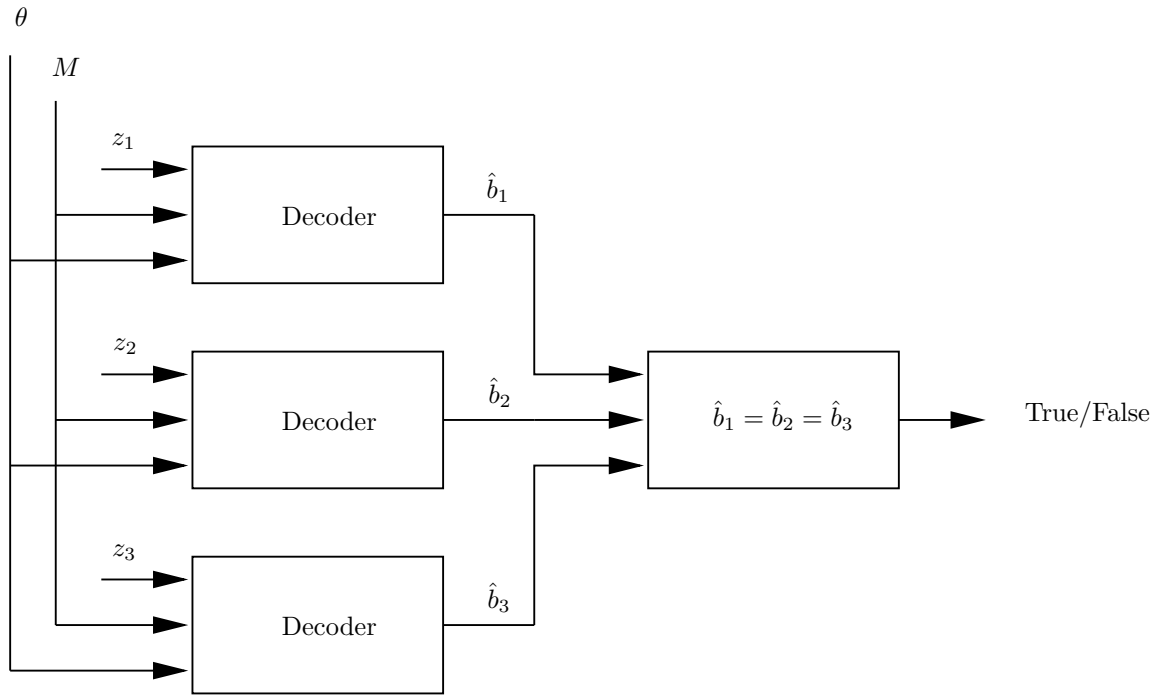

Figure 5. Block diagram of the exhaustive search of the secret key approach. We checked $1 \leq M \leq 25$ for increasing values of $\theta \in \mathbb{N}$, until $\hat{b}_1 = \hat{b}_2 = \hat{b}_3$.

3. EXHAUSTIVE SEARCH OF THE SECRET KEY

In a second stage of the contest, the watermarking algorithm was revealed: the well-known data-hiding method by Miller, Doërr and Cox.[6] This side-informed method is based on the use of trellis codes for performing the source and channel coding, being parameterized by some of the properties of the used trellis (number of states and number of arcs per state), as well as the spreading factor.

Still one open question for the attacker is how the original data-hiding scheme was adapted to this particular application, as a detection (i.e., zero-bit watermarking) scenario instead of a decoding (i.e., multiple bit watermarking) setup was being considered. Probably the most straightforward way of carrying out this adaptation is to compare the decoded message with a secret reference (which is the actual embedded message); if they are identical, the watermark is said to be present in the received signal, and absent otherwise. Trying to verify whether the system available in the BOWS webpage was based on these principles, and if the three images provided were watermarked with both the same key and the same reference message, we input those images to the detectors corresponding to the other two images, obtaining (of course) very small PSNRs, but confirming our intuition, as the watermark was still detectable.

Furthermore, we became aware of the existence of a new version of that algorithm, publicly available on the web of the authors.[7] In this implementation some values are assumed for the aforementioned parameters, in such a way that the decoder is only parameterized by the image to be checked $\mathbf{z}$, the secret key θ and the length (in bytes) M of the message to be decoded; this last point represents an additional difficulty to attackers' task, as they must also consider the different choices for that parameter. Despite these obstacles, we decided to peruse this implementation with the two-fold objective of better learning how it actually worked, and performing a security attack (here meaning an attack trying to gain knowledge about the chosen secret key).

Therefore, if we were able to find a pair (θ, M), such that the output of the decoding function provided by the authors of that trellis-based algorithm[6] were the same for the 3 provided images, we could be fairly confident that the used secret key was θ. Given that no a priori information about the value of the secret key was available,

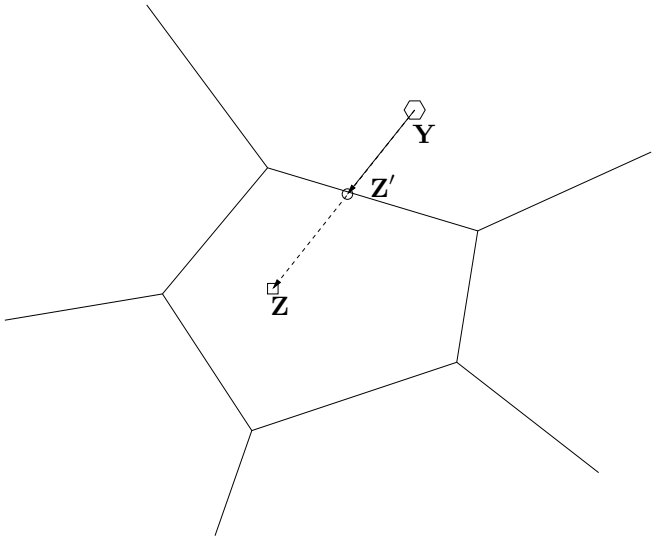

Figure 6. Diagram describing the generation of the attacked signal $\mathbf{Z}'$ from the watermarked signal $\mathbf{Y}$, and the signal $\mathbf{Z}$ obtained by embedding a message different of the reference one into the watermarked content.

we decided to try the exhaustive search mechanism. Furthermore, we had to establish a range of possible values for M, in order to try an exhaustive search for each possible value of the secret key. Taking into account that the studied scheme was really robust against most signal processing operations, it was clear that its rate (i.e., used coefficients per embedded bit) should be small enough. Considering that the number of coefficients of the provided images used by the algorithm for embedding information is $12 \cdot 512 \cdot 512/64 = 49152$, we decided that it was unlikely that the number of hidden bits were larger than 200, i.e. $M < 25$ bytes, since otherwise the number of coefficients per embedded bit should be smaller than 240, a choice which would not afford much robustness against conventional signal processing attacks. In view of these considerations, we decided to implement the exhaustive search plotted in Fig. 5. After 9 days of computation of a PERL script verifying if the decoded messages for different secret keys and lengths of the message were the same for all the three images, run on a shared computer with 2 Intel Pentium Xeon processors at 3.06 GHz and 2 GB RAM (be aware that the time requirement could be lightened if this process were paralleled) the authors found a pair of values (θ, M) verifying the condition introduced above.

Once the security of the system has been completely broken, the attacker has to devise how this information can be used to produce signals as close as possible to the provided watermarked signals, but where the watermark has been removed, i.e., where the decoded message is no longer the reference one. In order to do so, we have watermarked the provided signals with the same key θ, but with a different message of length M. In fact, taking into account the trellis nature of the used code, we only changed the last bit of the reference message, assuming that in that way the distance between the originally embedded codeword and the newly obtained one would be close to the minimum; this reasoning is based on the trellis structure of the codebook. Nevertheless, be aware that this strategy is not necessarily the optimal one, due to both the heuristic nature of the embedding algorithm, and the fact that a codeword of another message different of the slightly changed version of the reference message could be closer to the original codeword. In any case, computing the new signal in the described way, and considering the linear convex combinations of this signal and the originally provided one, we produced signals really close to the latter, but where the watermark has been removed; this procedure is depicted in Fig. 6. The PSNRs obtained for the 3 proposed images are, respectively, 53.5051 dB, 56.1106 dB and 55.6560 dB.

4. COMPARISON AND CONCLUSIONS

Clearly, the knowledge of the watermarking system was employed in our second approach to obtain results similar to those of the first one, but with a reduced computational cost. Nevertheless, we would like to remark that the

(a) (b) (c)

Figure 7. The three signals obtained using the attack based on the exhaustive search of the secret key.

first approach does not take into account *any* knowledge about the watermarking system, and still surprisingly good results are obtained, at the cost of a larger computation time. Furthermore, the complete disclosure of the secret was possible in this case due to its small length; in fact, an interesting conclusion to be drawn is that the space of the secret key should be large enough in order to prevent exhaustive search attacks. In this sense, the size of the space of the watermaked signal is not so important, as the attacker can focus his attack on a smaller space, as it is normally the space of the secret key. This is especially important, as the disclosure of the secret key does not only allow to obtain signals where the watermark is not detected with an impressive quality, but also allows the attacker to generate falsely watermarked signals (also known as *forgeries*), in all cases with almost no extra cost for successive contents, as the attack needs be performed once per secret key (not per content).

Another important conclusion regarding the BNSA is that it is possible to trade-off the final PSNR and the computational load; this compromise is achieved by reducing the number of gradient components that are actually computed.

Compared with the results obtained by other groups and recorded in the BOWS webpage,[1] we can see that the results achieved by our two attacks are only comparable with those obtained by Andreas Westfeld,[8] who achieved PSNRs of 60.74, 57.05 and 57.29 dB, respectively, in the stage of the contest where the watermarking method was publicly known. On the first stage, when that method was still not made public, the winner was the team led by Scott Craver, who achieved PSNRs of 39.67, 39.65 and 38.45 dB. These results clearly show the effectiveness of the proposed attacks.

REFERENCES

1. http://lci.det.unifi.it/BOWS/.
2. P. Comesaña, L. Pérez-Freire, and F. Pérez-González, "Blind Newton Sensitivity Attack," *IEE Proceedings on Information Security* **153**, pp. 115–125, September 2006.
3. J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, 1999.
4. D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, 1995.
5. M. Barni, F. Pérez-González, P. Comesaña, and G. Bartoli, "Putting reproducible signal processing into practice: a case study in watermarking," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'07*, (Honolulu, Hawai'i, USA), April 2007. Accepted.
6. M. L. Miller, G. J. Doërr, and I. J. Cox, "Applying informed coding and embedding to design a robust high-capacity watermark," *IEEE Transactions on Image Processing* **13**, pp. 792–807, June 2004.
7. http://www.adastral.ucl.ac.uk/ gwendoer/dptWatermarking/.
8. A. Westfeld, "Lessons from the BOWS Contest," in *Multimedia and Security Workshop 2006*, pp. 208–213, (Geneva, Switzerland), September 2006.